Narrating the "what" and "why" of our moral actions

Judy Sein Kim (judyseinkim@gmail.com)
Department of Psychology, Yale University
New Haven, CT 06511 USA

M.J. Crockett (mj.crockett@yale.edu)
Department of Psychology, Yale University
New Haven, CT 06511 USA

Abstract

To defend or burnish our moral reputation, we often tell moral narratives. Moral narratives describe morally relevant actions and explanations of those actions, detailing how people acted and why they did so. When and why does communication about moral events include descriptions of peoples' actions and explanations? We hypothesize that informational, reputational, and presentational goals of narrators shape whether their communication contains clear actions and explanations. We asked a group of "narrators" to communicate with other people following a moral decision. Another group of "audience" members judged them based on their chosen statement. We find that the informational and reputational goals of narrators can explain what information they decide to reveal. Narrators choose what to say based on inferring the audience's likely perceptions but underestimate how much audiences in fact expect answers to "what" and "why". Audiences, however, do not always perceive the lack of expected information as indicative of deceit.

Keywords: narratives; morality; pragmatics; language; decision-making

Introduction

We often rely on other people's words to learn about whether they acted morally. We may not have been present at the time that a moral event occurred, and we have no way of knowing what was going on inside someone's mind unless they tell us. This communication about moral events, people, and their motives often takes the form of narratives. Scholars from across many fields, ranging from Plato to Freud to Rumelhart to many more contemporary philosophers, linguists, narratologists, economists, historians, and psychologists have written about narratives (e.g., Rumelhart, 1975; Schank & Abelson, 1975; Zukier, 1986; Bruner, 1986; Green & Brock, 2000, McAdams, 2001).

The present work focuses on narratives with moral content. We build on the philosopher Greg Currie's definition of *narrativity* as a continuous characteristic of corpora (Currie, 2010). Corpora with high narrativity have two key elements. First, they contain descriptions of actions taken by particular people, whom we call *targets.* Second, they contain coherent explanations of the targets' actions. In other words, narratives answer two key questions under discussion (QUDs): "What did someone do?" and "Why did they do that?" (Roberts, 2012). Expanding on this framework, we define *moral narratives* as corpora that provide information about the moral character of one or more targets, typically by describing their morally relevant actions and explanations for those actions. Examples of moral narratives might include gossip about friends and co-workers, competing accounts of the same event in court, carefully crafted accounts in the media, and stories we tell ourselves about ourselves.

While past work has explored various phenomena related to moral narratives, several questions remain unanswered. Here, we propose a framework for understanding the function and construction of moral narratives in terms of the goals and beliefs of narrators and audiences. We hypothesize that for moral narratives to be satisfying for narrators and audiences, they need to answer the QUDs of "what" and "why". Exactly how narrators choose to answer those QUDs, however, depends on their specific goals and their beliefs about what audiences will infer from the narratives.

Constructing moral narratives

We build on the Rational Speech Act (RSA) model framework and understand narrative construction as choices in which speakers maximize utility according to particular goals (e.g., Goodman & Frank, 2016; Yoon & Tessler et al., 2020). Narrators consider a small number of possible moral narratives which have different values relative to each goal. Narrators decide on how to weigh each goal, resulting in a weighted sum (i.e., a utility) for each narrative. Moral narrative construction is then a process of selecting the narrative with the highest utility. In this paper, we aim to understand what kinds of goals narrators consider when telling moral narratives, as well as how they are valued according to each goal.

In our framework, moral narratives emerge in situations where a morally relevant event has happened and there is some need for communication *by* someone (the narrator) *about* someone (the target) *to* someone (the audience). For example, imagine that your roommate finds out their food has disappeared from the fridge and mentions it to you, expecting you to implicate or exculpate yourself in response. Here, we focus on such first-person narratives in which the "by someone – about someone – to someone" goal structure arises because an audience (e.g., your friend) expects a narrator to provide information about their own moral actions.

Once the expectation for a moral narrative is set up in a communicative exchange, narrators must consider an *informational goal*, or how informative they want to be for the audience (Grice, 1979). At a minimum, audiences expect narrators to describe the target's involvement (i.e., their actions) in relation to the triggering moral event. Additionally,

people often want to learn about why someone acted the way they did (i.e., an explanation) to make a more informed judgment about what kind of person they are (Malle, 2021). In other words, to be cooperative and informative, narrators must accurately infer that the QUDs of the exchange ask for reports of actions and explanations, as well as then deciding to prioritize the informational goal. Further, to be truly informative, narrators must do more than provide the right kinds of expected information: they must also communicate what they believe to be true (i.e., their subjective, believed truth).

However, informational goals alone cannot account for the range of ways in which people communicate within the same goal structure. A rich literature on deceptive communication suggests that people often choose to be uninformative. For example, people often provide no information at all, provide factually truthful but irrelevant and misleading facts (i.e., palter), be vague or ambiguous, or evade the topic (e.g., Clementson, 2020; Rogers et al., 2019). These tactics emerge when narrators want to hide the truth because they have some goal that conflicts with their desire to tell the truth. For example, narrators may want to make the narrative's target seem like a better or worse person than they actually are, at the expense of the truth.

One way to achieve such a *reputational goal* is to describe false actions and explanations that portray the target in the desired direction (positive or negative). Narrators can also pursue reputational goals by withholding relevant information. If the full narrative truth is undesirable, even a vague or incomplete narrative that is, for example, unclear or missing details about actions and explanations, may better serve the narrator's reputational goal than the truth. These two strategies reflect different balancing of informational and reputational goals: while lying throws truth out the window, it can clearly direct audiences' beliefs about the target's moral character; and while vague narratives maybe be neither informatively nor reputationally ideal, they do not entirely sacrifice one goal for the other. Our framework therefore suggests that the balancing of these informational and reputational goals predict not only when moral narratives contain actions and explanations, but when they do not, as well.

Finally, narrators may also consider audiences' inferences about their balancing of informational and reputational goals (i.e., a second-order *presentational* value, like in Yoon & Tessler et al., 2020). Especially in cases where a narrator's informational and reputational goals trade-off (e.g., wanting to seem morally better than they are), the narrator may want to mask their dishonesty. Missing information might cue audiences towards reputational goals that are at odds with informativity. For example, an incomplete narrative lacking clear and relevant descriptions of actions or explanations may arouse suspicion of intentional deception. Therefore, we hypothesize that narrators' inferences about audience expectations of information are important for how narrators compute the presentational value of moral narratives.

In summary, we propose that the core of moral narrative construction involves the balancing of all of these goals: to be cooperative and informative (informational goal), to sway audiences to view the target's reputation in a certain light (reputational goal), and to appear like an honest and cooperative communicator (presentational goal).

Key Questions and Predictions

Across two experiments involving narrators, audiences, and pre-written statements that narrators could select, we test and explore some of the key assumptions and predictions of the theory. The results are presented as four components:

Audience expectations of information: First, we validate that our experiment sets up the kind of context in which we would expect canonical moral narratives. Following a moral event, narrators are expected to communicate about their own moral actions to an audience. We examine whether given this set up, audiences indeed expect clear description of actions and explanations from narrators.

How narrators value different narratives: We then explore the informational, reputational, and presentational values of each narrative by asking narrators questions about their own beliefs as well as about the audiences' likely perceptions. For informational value, we ask how much narrators agree with each statement. For reputational value, we examine inferences narrators make about audiences' likely moral judgments upon reading their narrative. For presentational value, we look at narrators' inferences about audience expectations of actions and explanations.

Whether narrators' goals explain their narrative choices: Crucially, we hypothesize that considering how narrators weigh all these goals can predict narrative choices better than any one goal alone. We predict that narrators' balancing of competing goals can explain when narrators provide incomplete details about actions and explanations, thereby compromising informativity.

Audience inferences about missing actions and explanations: Finally, to understand whether the missing information is perceived as suspicious, we explore inferences audiences make about the causes of incompleteness (e.g., whether it implies a competing reputational goal).

Overview of Experiments

In the Narrator Experiment, "narrators" made a moral decision to claim a bonus at the risk of a co-worker losing their earnings. They were told about an audience who would be deciding whether to increase or decrease their bonus based on their statement. Narrators were told that audiences would not know whether they claimed the bonus and risked their co-worker's earnings. In the Audience Experiment, a different group of participants ("audience members") were told about the narrators, read statements that narrators chose to show

them, and answered questions about the narrators and their statements.

Narrators were presented with a small number of narrative choices which varied in how much information they contain, in a 2x2 design where an action and explanation were directly described or not described (vague, unrelated statement) (Figure 1). A full statement that contained action and explanation (AE, action + explanation) stated that "I claimed the raise," and that "I really didn't care about what happened to the co-worker". For vague action and explanation (VV, vague action + vague explanation), the choice was "This was a weird experiment. I would be interested to see the final results of the study". The AV narrative contained the clear action and vague explanation, and VE narrative the vague action and clear explanation.

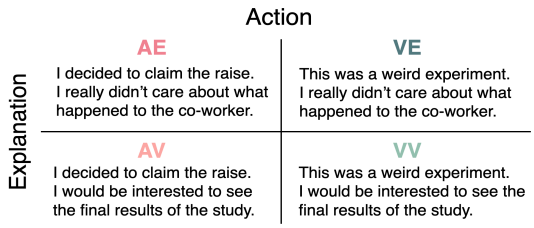

Figure 1: Narrative choices

In the Audience Experiment, audiences were told about the choices narrators had to make and given all the instructions that the narrators saw. Audiences first gave free response answers about what kinds of information they were expecting from narrators. Next, audiences read each of the 4 narratives without being told whether each narrator claimed the raise or not. To test whether narrators' inferences about audiences are correct, we asked audiences questions that parallel those narrators answered about reputational and presentational values. Audiences were asked to adjust the narrators' bonus (reputational value) and judged whether each narrative contained all the information they were expecting. (Note that from a narrator's perspective, how a narrative presents an image of informativity is important for its presentational value, but from the audience's perspective, whether a narrative contains expected information may be a cue to the narrator's informational goal). To ask whether audiences interpret a low priority on information as suggestive of a high competing reputational goal, audiences were further asked about how they thought narrators were balancing informational and reputational goals (i.e., presentational goal), and why the narrators might be communicating that way (free responses, as well as ratings about questions about possible guilt, deceit).

Narrator Experiment

Methods

Participants. We recruited 150 participants from Prolific to be "narrators". Of these, 9 participants were excluded from analysis for failing attention checks or writing nonsense responses in free-response sections. As described below, participants were asked to make a decision affecting the earnings of another person. We designed the incentives of the decision task to prompt most participants to choose the "selfish" option, and accordingly, 100/141 did so. Only these participants went on to complete the rest of the experiment.

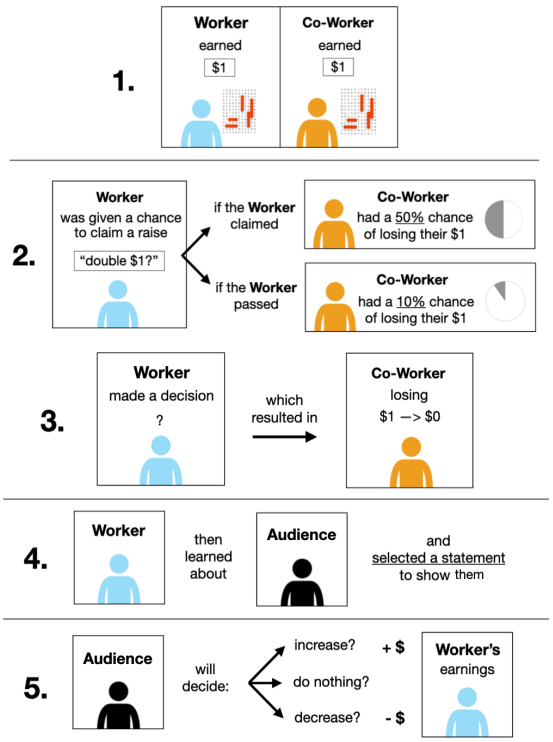

Figure 2: Schematic of experimental design

Experimental design and procedure. Figure 2 gives an overview of the experimental procedure. In the first part of the study, participants solved a simple word search puzzle. For completing the puzzle, narrators were told they earned $1. Narrators then were given a choice of "claiming a raise" to their puzzle earnings, which would double their $1 to $2. Narrators were told the choice changed a risk imposed on the earnings of a co-worker. If they declined to claim the raise, the co-worker has a 10% chance of losing all of their earnings, but if the participant claimed the raise, the risk would increase to 50%. Narrators were told that their choice would remain a secret to the co-worker.

All narrators included in this analysis decided to claim the raise. After making the choice, narrators found out that the co-worker ended up losing their earnings. They were then informed about the existence of an "audience" and asked to choose 1 of 4 statements (Fig. 1) to show them. Next, narrators saw each of the narratives one-by-one and answered detailed questions on sliders ('definitely no – not sure – definitely yes'), including what they thought audiences would think upon seeing that narrative. To obtain continuous measures of narrators' overall preferences for each narrative, we asked "Are you likely to **show** this statement to the audience"? For informational value, we asked, "Do you **agree** with this statement?". For inferences about audience

perceptions of informativity, we asked "Based on reading this statement, do you think the audience will be **satisfied** that you gave them **all the information they wanted**?" Finally, for inferred reputation, we asked, "Based on reading this statement, how do you think the audience will **adjust** your earnings?" which participants could respond on a scale of 'Give extra $1 - no change - Take away $1'.

Audience Experiment

Methods

Participants. An additional 100 participants from Prolific were recruited to be members of an "audience". Of these, 7 participants were excluded from analysis for failing attention checks or writing nonsense responses in free-response sections, resulting in 93 participants in this analysis.

Experimental design and procedures. "Audience members" in this experiment were told that they would be reading about what other workers did during another study. They saw many of the same instructions as the narrators, but from a 3rd person perspective. Audiences knew that narrators had solved a puzzle, received a bonus, then given a choice to claim a raise that affected the earnings of a co-worker. Like the narrators, audiences found out that the narrator's co-worker ended up losing their $1. Audiences, however, were not told whether the narrator had claimed or not.

Audiences were then told that narrators selected a statement to show them. Audiences were presented with 4 different narrator's statements (corresponding to the 4 narratives, AE, AV, VE, and VV). Audiences then answered questions analogous to those from the previous experiment. Importantly, audiences were asked to adjust the narrators' bonus: "Would you like to **adjust** the Worker's bonus (increase or decrease any amount between $0-$1)?". Audiences were also asked whether the narrative provided information they were expecting: "Is this **all the information** you wanted to know?"

Results

Audiences expect both actions and explanations

Audiences were asked, "What kind of information do you want from the [narrator's] statement in order to decide whether to adjust their earnings?" Free response answers revealed 48% of audiences wanted to know "Did you claim?" and 32% wanted to know, simply "Why?". 28% of narrators additionally wanted to know answers to specific "Why" questions, such as how the narrator felt toward the co-worker (concern or malice), and whether they deserved or needed the money.

When asked to judge whether the 4 narratives provided all the information they expected from narrators, audiences judged AE to be highest in information satisfaction (Fig. 3 left panel, ratings ranging from 0 to 100, M=88.81), followed by AV (M=71.82), then VE (M=51.1), then VV (M=16.37). A mixed effects model confirmed that actions and explanations were both expected, although actions more so than explanations (effect of action: β=55.45, SE=3.77, p<0.001; effect of explanation: β=34.73, SE=3.46, p<0.001). Further, there was an interaction effect such that once the action is described, additionally including the explanation does not increase satisfaction as much as it does to add an explanation when no action is included (interaction: β=-17.74, SE=4.51, p<0.001).

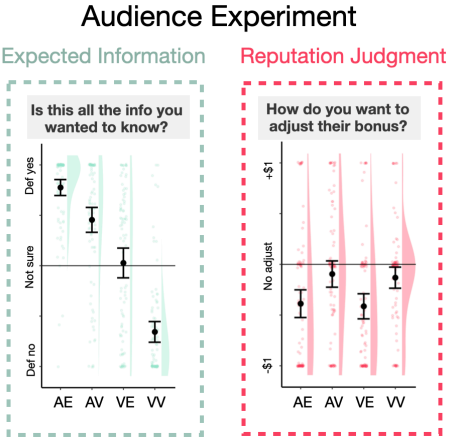

Figure 3: Audience expectations of information and judgments about narrators' reputation (whether to adjust their bonus). Error bars are 95% CIs.

Actions and explanations impact informational, reputational, and presentational values of narratives

Informational value. Narrators agreed most with AV (M=88.31), followed by VV (M=75.17), then AE (M=49.86) and VE (M=42.85) (Fig. 4; right panel). Narrators agreed with statements slightly more if the action was included but agreed a lot less if the explanation was included (effect of action β=13.14, SE=3.6, p<0.001; effect of explanation β=-32.32, SE=3.6, p<0.001). There was a small interaction where the explanation combined with the action (AE) lowered agreement beyond the negative effect of explanation alone (β=-6.13, SE=5.09, p=0.02)

Reputational value. Narrators inferred that audiences might reward the vague narrative (VV, M=55.08), take a little bit of money away for admitting the action (AV, M=45.72), but take away a lot of money for giving the selfish explanation (VE, M=17.15), and even more away for admitting both the action and explanation (AE, M=10.72) (Fig. 4; middle panel). In other words, narrators thought that both actions and explanations would decrease reputational value, with explanation having a very strongly negative effect (effect of action: β=-9.36, SE=2.41, p<0.001; effect of explanation: β=-37.93, SE=2.63, p<0.001, no interaction).

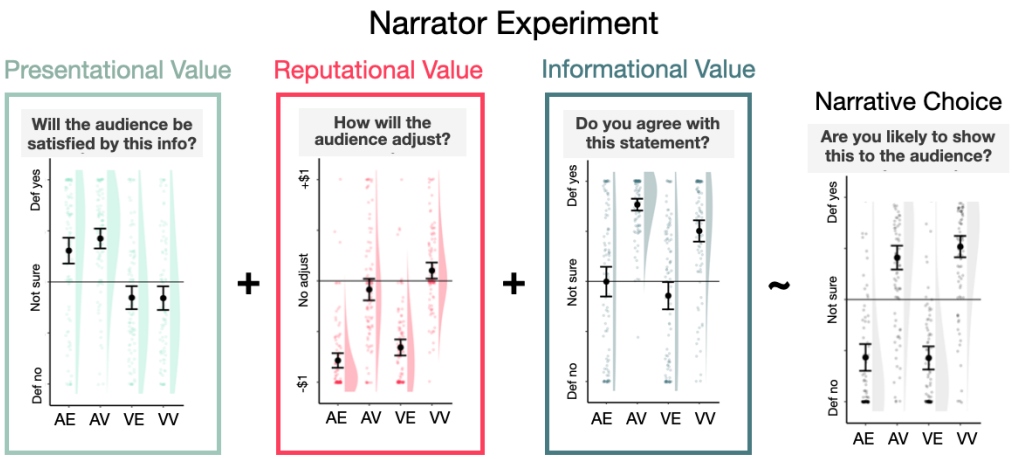

Figure 4: Results from Narrators Experiment. Narrative choices and answers to questions about informational, reputational, and presentational values. Error bars are 95% CIs.

Presentational value. When asked to guess how satisfied audiences will be about the information contained in each narrative, narrators guessed that audiences would be satisfied as long as the narrative contained actions and that the inclusion of explanations would not affect satisfaction either way (Fig. 4; left panel, M=65.31 for AE and M=71.24 for AV, followed by M=42.48 for VE and M=42.05 for VV; effect of action: β=29.19, SE=3.31, p<0.001; no effect of explanation or interaction).

Informational, reputational, and presentational goals predict narrators' choices

To index narrative choice, narrators were asked how likely they were to show each narrative to the audience (Fig. 4, far right). Preference was high for VV (M=75.87) and AV (M=70.5), followed by AE (M=21.65) and VE (M=21.4). These choices were most affected by the presence of the explanation, where the narrators indicated they were less likely to show the AE or VE (effect of explanation: β=-37.93, SE=2.63, p<0.001; no effect of action or interaction).

To ask whether balancing multiple goals (i.e., informational, reputational, and presentational goals) best predicts narrative choice, we compared multiple linear mixed-effects models with informational, reputational, and presentational values as predictors for whether narrators were likely to show the narrative.

The full model with informational ("agree"), reputational ("how will the audience adjust your bonus"), and presentational ("will the audience be satisfied") values included (r^2=0.66) revealed that reputational value contributed most to narrators' narrative choices (β=0.9, SE=0.05, p<0.001), followed by informational value (β=0.34, SE=0.04, p<0.001), then more weakly, presentational value (β=0.09, SE=0.04, p=0.04). We compared models using Bayes Factors and found that this model accounts for the narrative choices better than reputational only (log BF=-22.83) and informational only models (log BF=-83.96). However, the informational and reputational only model performed slightly better than the three-goal model (log BF=3.18).

Together, this suggests that narrators show the AV and VV narratives most frequently because they achieve a balance of informational and reputational goals, which narrators in this experiment prioritized. AV has high informational value and is not too low on reputational value relative to the other choices, and VV has high informational and reputational values. Since narrators cared most about reputational value in this experiment, they avoided the narratives that contained the selfish, reputationally costly narratives. Further, while VV was presentationally low in value, narrators still preferred it given their low prioritization of presentational goals.

Audience perceptions do not match narrators' inferences

Audience judgments about narrators' reputation. While narrators thought that the vague narrative (VV) may be looked upon most favorably by the audience (Fig. 4 middle panel), audiences' actual decisions about whether to increase or decrease narrators' bonuses revealed that narrators who showed VV were not rewarded above any other narrative (VV=43.44, VE=29.32, AV=45.22, AE=30.61, Fig. 3 right panel). Audiences on average decided to take away money for all narrators, and more so for narrators who gave the selfish explanation (effect of explanation: β=-14.12, SE=3.41, p<0.001, no effect of action or interaction).

Audience perception of missing actions and explanations. As presented in Fig. 3 previously, audience expectations of information also diverged from narrators' inferences about audience informational satisfaction (Fig. 4, left panel). To investigate whether audiences find missing information suspicious (i.e., suspect a reputational goal that conflicts with informational goals), we asked audiences to guess why narrators were communicating the way they were ("What do

you think is the reason why the narrator is (or is not) communicating frankly and clearly?").

This revealed that on the one hand, audiences do most frequently suspect deception (e.g., "they're trying to mislead me" or "they're hiding something") for the VV narrative (Figure 5). However, audiences also do not always suspect vagueness as indicative of a nefarious goal. They also cite a variety of different reasons why narrators may be vague, such as incompetence ("I think they're just bad at communicating"), laziness ("they didn't care enough"), as well as feelings of shame or guilt ("they felt bad about what they did").

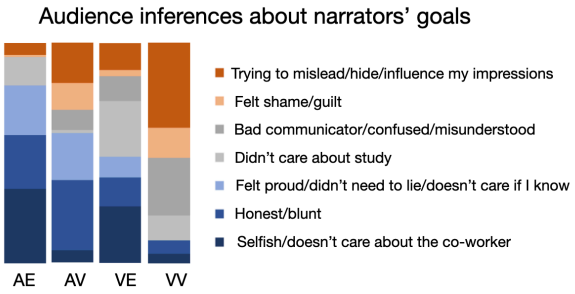

Figure 5: Audience free response reasons why the narrator is communicating the way they are

Discussion

Here, we introduce a framework for understanding moral narrative construction as driven by narrators' communicative and social goals. Using experiments with narrators and audiences, we tested the idea that narrators' informational, reputational, and presentational goals can predict whether moral narratives contain details about the moral actions and explanations of a target. We focused in particular on understanding when communication about moral events sometimes does *not* contain answers to "what" and "why" questions under discussion.

We find that in deciding what to say, narrators factor in the truth of a statement (informational value), as well as an audience's likely perception of their moral character (reputational value), and to a lesser degree, whether the audience will be satisfied by the information (presentational value). Informational and reputational goals together predict narrators' choices better than each of the goals alone. In the current experiments, these goals predicted whether narrators included clear descriptions of what they did (claimed a raise in earnings, imposing a risk to a co-worker) and why ("I didn't care about the co-worker"), as opposed to vague, irrelevant statements.

An important part of our framework is understanding when and why moral narratives are told. We focused on one type of goal structure in which the audience expects information from the narrator. Audience members in this experiment clearly expected descriptions of actions and explanations. Interestingly, however, narrators predicted that audiences would only want to know about their actions and would not experience increased satisfaction in hearing the explanation for their action. This inference is particularly puzzling because explanations strongly affected how narrators answered questions about the informational and reputational values of narratives, as well as, ultimately, their narrative choices. In other words, the inclusion of an explanation for their action was clearly important for narrators—so why did they incorrectly infer that audiences would not similarly find it important?

One possibility is that narrators in this experiment were simply bad at perspective taking or inferring others' perceptions. Narrators were, after all, also wrong in thinking that the vague narrative would be rewarded by the audience. Another explanation is that narrators' errors may reflect some self-serving biases (e.g., Stanley, et al., 2017). For example, narrators may have chosen to believe that audiences do not expect explanations because they were worried about audience judgments. One future direction is to compare narrators' inferences about audience perception when there are no social or material consequences.

An important limitation of this study is that people in real life do not decide what to say by selecting among a small number of choices given to them. Further, we chose the full narrative (AE) to pit informational and reputational goals against each other, with a trade-off that many narrators did not find it reflective of their own reasons for deciding. We expect that when informational and reputational goals can be aligned within a narrative that contains an action and explanation, people will choose more complete narratives. Further, there are ways of providing the right kinds of information while neither lying nor telling the full truth. For example, paltering involves "not false" but misleading statements (Schauer & Zeckhauser, 2007; Rogers et al., 2017). The choices in this experiment were limited, but we hypothesize that narrators' goals as studied here will be able to account for other variations in moral narratives.

Finally, why did narrators not weight presentational value more? We hypothesized that audiences suspect reputational goals conflict with informativity when they detect missing information (i.e., actions or explanations). One possible reason narrators in the current experiment were not too concerned with presentation is that they correctly guessed audiences would not necessarily suspect deceit. When narrators presented the vague narrative, audiences were often understanding, guessing that a narrator may be unable to clearly describe their actions and explanations due to lack of communication skills, laziness, or interfering emotions. Put differently, not everyone can tell perfect narratives, and neither do people expect them to do so. While we demonstrate that narrator goals are predictive of narrativity, learning, experience, and ability are also key factors. This raises interesting questions about how people become "good" narrators, as well as what functions are served by statements with high narrativity. Further, we predict that in other contexts, such as when the stakes of being found lying or not being believed is higher, narrators will weight presentational value as more important.

Acknowledgments

This work is funded by the John Templeton Foundation grant on How Stories Inspire Virtuous Behavior awarded to MC.

References

Bruner, J. (2004). Life as narrative. *Social research: An international quarterly*, *71*(3), 691-710.

Buller, D. B., Burgoon, J. K., Daly, J. A., & Wiemann, J. M. (1994). Deception: Strategic and nonstrategic communication. *Strategic interpersonal communication*, 191-223.

Clementson, D. E. (2018). Deceptively dodging questions: A theoretical note on issues of perception and detection. *Discourse & Communication*, *12*(5), 478-496.

Currie, M. (2010). *Postmodern narrative theory*. Macmillan International Higher Education.

Gaspar, J. P., Methasani, R., & Schweitzer, M. (2019). Fifty shades of deception: Characteristics and consequences of lying in negotiations. *Academy of Management Perspectives*, *33*(1), 62-81.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of personality and social psychology*, *79*(5), 701.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41-58). Brill.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, *20*(11), 818-829.

Kao, J., Bergen, L., & Goodman, N. (2014). Formalizing the pragmatics of metaphor understanding. In *Proceedings of the annual meeting of the Cognitive Science Society* (Vol. 36, No. 36).

Mahon, J. E. (2008). The definition of lying and deception.

Malle, B. F. (2021). Moral judgments. *Annual Review of Psychology*, *72*, 293-318.

McAdams, D. P. (2001). The psychology of life stories. *Review of general psychology*, *5*(2), 100-122.

Ransom, K., Voorspoels, W., Perfors, A., & Navarro, D. (2017). A cognitive analysis of deception without lying. Cognitive Science Society.

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and pragmatics*, *5*, 6-1.

Rogers, T., Zeckhauser, R., Gino, F., Norton, M. I., & Schweitzer, M. E. (2017). Artful paltering: The risks and rewards of using truthful statements to mislead others. *Journal of personality and social psychology*, *112*(3), 456.

Rumelhart, D. E. (1975). Notes on a schema for stories. In *Representation and understanding* (pp. 211-236). Morgan Kaufmann.

Saul, J. M. (2012). *Lying, misleading, and what is said: An exploration in philosophy of language and in ethics*. Oxford University Press.

Schank, R. C., & Abelson, R. P. (1975, September). Scripts, plans, and knowledge. In *IJCAI* (Vol. 75, pp. 151-157).

Stanley, M. L., Henne, P., Iyengar, V., Sinnott-Armstrong, W., & De Brigard, F. (2017). I'm not the person I used to be: The self and autobiographical memories of immoral actions. *Journal of Experimental Psychology: General*, *146*(6), 884.

Wilson, D., & Sperber, D. (2002). Relevance theory.

Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2020). Polite speech emerges from competing social goals. *Open Mind*, *4*, 71-87.

Zukier, H. (1986). The paradigmatic and narrative modes in goal-guided inference.